DOCUMENT RESUME

ED 109 173 TM 004 620

AUTHOR Kocher, A. Thel
TITLE Issues in the Formative Evaluation of Alternative Schools.
PUB DATE [Apr 75]
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D. C., March 30-April 3, 1975); Not Available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.76 PLUS POSTAGE. HC Not Available from EDRS.
DESCRIPTORS *Alternative Schools; *Decision Making; Evaluation Methods; *Formative Evaluation; Models; Summative Evaluation

ABSTRACT

The paper examines many of the evaluation problems noted by the Phi Delta Kappa National Study Committee on Evaluation and others and sets forth ideas about the practice of formative evaluation to avoid those conflicts and provide educational decision-makers with useful information. The paper also discusses a decision-making model conceptualized by one alternative school's evaluation group and how the evaluator's role relates to serving decision makers under the model. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED109173

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIG N
ATING IT POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

ISSUES IN THE FORMATIVE EVALUATION OF ALTERNATIVE SCHOOLS

A position paper by

A. Thel Kocher
Minneapolis Public Schools
Southeast Alternatives Internal Evaluation Team

Presented at the annual meeting of the
American Educational Research Association
Washington, D.C.

April, 1975

BEST COPY AVAILABLE

TM 004 620

Preface

Many of my thoughts about evaluation which are expressed in this paper have been influenced by my interactions with my colleagues in the Southeast Alternatives Project. To them, especially to those on the Internal Evaluation Team, I wish to express my gratitude. Of course, I alone am responsible for any errors of omission or commission.

A special thanks to Susie Demet for her editorial assistance and to Rose Salo and Lois Caswell for their clerical assistance.

Introduction

I believe that educational evaluation as practiced in the majority of instances has failed to live up to peoples' expectations. It has failed to give practitioners the information they needed to improve education. Educators, especially those involved in developing alternative programs, have come to view evaluation as a process having little value to them; indeed, in many cases, they view it as antithetical to their purposes.

Can evaluation be improved to make it a worthwhile endeavor? Participants in the Southeast Alternatives project do believe that evaluation is worthwhile--that it does provide them with useful information. Evaluators in the project believe that much of the commonly encountered negativism toward evaluation is the result of misadvice and misunderstanding arising out of the kind of evaluator-practitioner relationships which were noted by the Phi Delta Kappa National Study Committee on Evaluation (Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman & Provus, 1971) when they said:

> Professional evaluators can no longer afford to give the practitioner the cavalier, arrogant and condescending treatment that has so often characterized their relationships in the past (p. 4).

SEA evaluators believe that by working <u>with</u> decision makers rather than treating them in the way the committee has noted we can greatly reduce or eliminate this negativism. The present paper sets forth many of the SEA views about formative evaluation and how we believe it should be practiced in order to provide decision makers with useful information. While the paper is focused specifically on the formative evaluation of alternative programs most of the suggestions presented seem equally applicable to summative efforts.

The Role of Internal Evaluation

First of all, the question of what purposes evaluation of alternative programs must and can serve should be considered. Mike Hickey (1972) has stated four specific purposes of evaluation in alternative programs:

> First, and perhaps of highest priority, is the purpose of internal self-improvement for the program, which in turn relates to the ongoing planning process (informal though it may be).
>
> Second, as a basis for establishing the credibility of the alternative program, evaluation must meet the demands of a variety of "publics". . . .
>
> Third, a primary rationale for the existence of alternatives within public education is that they become the means or the process by which public education evolves. Realistically, some strategies for educational alternatives will not work. Evaluation provides a base for identifying those that work and those that don't. . . .
>
> Finally, the evaluation of student progress is difficult without an adequate understanding of where the program itself stands (p. 2).

The National Institute of Education, apparently recognizing the value of the formative evaluation which Hickey refers to in his first goal, has provided within the Experimental Schools Program a local project evaluation component. According to the NIE this local component

> provides for a legitimate, internalized role of evaluation, and can be tailored specifically to an individual project and its goals. It should provide for quick feedback and enhance resident and staff participation. It is designed to aid in meeting objectives and improving performance. (Memorandum from R. Binswanger to J. Kent, April 12, 1974).

In order for the local SEA evaluation component to carry out that charge from the NIE, nearly all evaluation activities the component conducts are of a formative nature. Formative evaluation of this type may be characterized as follows:

- its purpose is to improve the program;
- its primary audience is always the program participants--students, parents and school personnel;
- different evaluation situations dictate the conditions of reporting--to whom, report medium, type of recommendations, etc; and
- the evaluators must be responsive to the needs of the program personnel.

Hickey's other three purposes seem to call for the evaluation information used in a more summative sense. Certainly, these are important but it is the primary purpose of the present paper to discuss ways of successfully implementing a formative effort that can meet the information needs of decision makers.

Providing Effective Service

It seems that one of the major reasons for the failure of evaluation to provide decision makers with the information they need to make education better has been the lack of adequate communication between evaluators and decision makers. Not simply failure of evaluators to effectively communicate results but their failure to involve decision makers at every step of the process from deciding what questions should be asked to interpreting the results.

A recent study (Lange & Speiss, 1974) has shown that of four factors related to communication between evaluator and decision maker--frequency of interaction, interpretability of reports, advocacy for his conclusions, and credibility of the evaluator--credibility is the most important factor in determining whether or not practitioners will

use the results of an evaluation study. The SEA evaluation component has long believed in an approach that stresses credibility.

Many of the past short-comings of evaluation have led practitioners to view evaluators as uncredible. If evaluators are to overcome this image, they must be aware of these past short-comings and of possible ways to eliminate them. I want to note what I believe are many of the reasons for this lack of credibility and what SEA evaluators believe might be done to remedy the situation.

Past Shortcomings and Hopeful Remedies

Perhaps part of the negativism with which practitioners have viewed evaluation has been due to the heavy emphasis in program evaluation on summative efforts. Formative evaluation efforts designed to provide information to assist in making ongoing changes have, in the past, been centered primarily in the areas of student evaluation and curriculum (product) development. Only recently have there seemingly been significant attempts to extend these formative efforts to program evaluations in order to provide practitioners with information which can be utilized to improve the program as it evolves. Past summative efforts with feedback coming at the end of a project and with emphasis on providing only an overall assessment of effectiveness have failed to provide this needed feedback. These summative evaluations have been classified as "good" only when they developed no negative information. In contrast, in formative evaluation efforts negative data is usually more productive than positive data since it points to areas of the program which have weaknesses at a time while it is still possible to make program modifications in an attempt to eliminate these weaknesses.

It seems that both decision makers and evaluators must share the blame for the lack of formative evaluation efforts in the past but now both groups must work together to insure that some type of internal formative evaluation effort is included as an integral part of every program development effort. Evaluation can, when properly practiced in a formative sense, provide the answers to many of the decision makers questions; to avoid it because it has failed in the past is a "cop-out."

Another major reason for much of the negativism toward evaluation may very well be due to a lack of understanding of what evaluation is all about. Mike Hickey (1972) has noted this as follows:

> much of the hostility toward evaluation has arisen from failure to understand what evaluation is all about. The blame for this failure rests equally with both sides: the evaluator for not clarifying what he is doing and why, and for imposing an evaluation design on the program; the program director for not requiring full explanation of the process and its underlying rationale and for not taking the initive for developing, at least partially, the evaluation design (p. 1).

Probably largely as a result of this failure of project personnel to demand an explanation of the process, much of what has been passing as evaluation has not been evaluation but has, rather, been research. Research methodology and evaluation methodclogy are not the same and attempts to equate

> them forces certain constraints inimical to the purposes of evaluation and makes it impossible to meet certain of the needs served by good evaluation (Stufflebeam et al., 1971, p. 22).

The purpose of research is to produce new knowledge that is generalizable and classic research design procedures call for experimental control of extraneous variables to accomplish this goal. In evaluation the goal should be to delineate, obtain and provide information for decision making. This information is not necessarily new knowledge and

is highly specific to a given situation. The decision maker needs data on how well his/her goals are being achieved in the real life situation, not on what would happen in a "laboratory" situation with interference from extraneous variables eliminated.

As there are differences in the design of research and evaluation studies there are also different requirements as far as the kinds of measurement instruments and observation techniques that are needed. As Rosen (1973) has noted:

> Experimental research always requires valid and reliable measurement instruments. Validity and reliability are also important criteria for evaluation instruments, but as the purpose of evaluation is to improve an enterprise, the evaluator cannot limit herself to measuring only variables for which there are tested instruments and techniques (p. 4).

Certainly, SEA evaluators believe that validity is important--in fact, one of the major problems in using only instruments and techniques that have been shown to be reliable is that they do not produce valid answers to many of the questions important to decision makers. However, evaluators should not be expected to conduct statistical validity studies on every instrument or technique they develop. Face validity assessments must suffice.

Also, as evaluators develop instruments or techniques they must follow sound practice to achieve devices that will produce reliable information but, again, they must not be expected to conduct reliability studies for every device.

Deemphasizing the importance of reliability and statistical validity seems especially appropriate to formative program evaluation efforts. As opposed to many summative evaluations, irrevocable decisions are not being made in which the life of the program is at stake. The formative information should be used to assist the decision maker in determining

needed program changes. If these changes are made and do not "pan-out", possibly as a result of unreliability or invalidity of information, continual formative monitoring will allow corrective action to be taken quickly.

Attempts to deal with evaluation as if it were research has also lead to a tendency to cast evaluation studies in the classic experimental/control group mold so common in research. This has been done in spite of the fact that as early as 1963 Cronbach stated that: "The aim to compare one course with another should not dominate plans for evaluation (p. 13)." If evaluation information is to be useful to the decision maker s/he should be concerned with how the program is meeting its goals, not how it compares with some competing program. The important thing is for the evaluator to look at the teaching/learning process and to provide information to enable the decision maker to strengthen weak aspects of the program.

Many of the problems which arise from the failure to recognize the distinctly different purposes of research and evaluation studies may well stem from the training of most early evaluators. When the need for a large number of evaluators first arose the majority of those drawn into the field had strong roots in the areas of psychological and educational research. They were steeped in the methods of classical research and tried to "bend" evaluations to fit these methods rather than to adopt new more appropriate methods. In the past decade, particularly in the past five years, enough has been written of the differences; but, evaluators trained in appropriate approaches are still not readily available. Practitioners must demand evaluators who are evaluators not researchers and they will get them.

Another reason why evaluators often fail to answer many of the questions that are important to decision makers has been stated as follows:

> With their attention riveted on achievement, evaluators overlooked the process because the really important data, product data, would be available only at the end, when student behaviors could be checked against the objectives (Stufflebeam et al., 1971, p. 13).

This failure of evaluators to consider the process has been inexorably linked to their insistence that the evaluation process be centered around previously defined behavioral objectives for the program. To be sure, goals and objectives are important but evaluators must come to realize that not everything can properly be assessed in terms of student product data. Particularly, in formative evaluation efforts evaluators must consider the process as more important than the product.

Another misunderstanding that has lead to many problems has been the failure of evaluators to discuss with practitioners the different ways in which evaluation may be defined, to make them aware of the consequences of the various definitions and for evaluators and practitioners to jointly agree on which definition should be adopted for a particular task. Two common definitions (Stufflebeam et al., 1971 & Gephart, 1973) and their implications are as follows:

Evaluation is identical to measurement. This approach involves administering a written norm-referenced test and interpreting the scores as the evaluation. This provides a seemingly objective way to evaluate because the scores are compared to a norm. But, this obscures the fact that, eventually, somewhere in program decision making value judgments are necessary. Since this method relies only on written norm-referenced tests, it has resulted in evaluations that have been narrow in focus.

It has not dealt effectively with evaluation of higher order cognitive and affective learning. With its emphasis on outcome measures it has also failed to provide evaluation of the teaching/learning process. Evaluation of the process is especially important if evaluation efforts are to provide the feedback necessary to assist in the ongoing program development. Evaluators can and should use a variety of methods for gathering information--tests (mostly objectives based), surveys, interviews, observations, etc. Many of these methods are now shrugged off because they lack reliability; but, as noted earlier, this should be considered grounds for exclusion. It goes without saying that not all of these methods are appropriate in a given situation but they are all tools which should be considered for appropriateness.

Evaluation equals judgment. In this approach an expert is asked to judge the program. The drawbacks of the approach are many. The evaluator is making impressionistic judgments with the data and the criteria used to assess them hidden and it is difficult to determine the validity of the judgments. The expert often does not systematically consider the decision maker's goals so the data is frequently irrelevant. Most importantly, since this type of evaluation presents judgments rather than data to the practitioner it takes the decision making process (at least much of it) out of the hands of the practitioner and places it in the hands of the evaluator.

The evaluation is identical to measurement and the evaluation is judgment approaches used alone, seem to have unsurmountable validity problems, particularly for alternative schools in which major emphasis is placed upon more than low order cognitive learning and in decentralizing the decision making process. In the SEA project the main purpose of internal evaluation is to assist decision makers. Thus, project in-

ternal evaluation efforts fall under a final definition given by Gephart (1973).

Evaluation serves decision making. Under this definition evaluation is considered as an integral part of the decision making process. The specific decision to be made gives direction to the evaluation effort and evaluation provides information to assist in making the decision but it does not make the decision.

A promising mode! Operating under Gephart's definition SEA evaluators believe decision making/evaluation should follow a process similar to the following (Rawitsch, 1974):

> Given: We must make a decision. To make this decision, we need the answers to certain questions. To get the answers to these questions, we need certain information. To get this information, we need "data" gathering strategies. (p. 4).

Thus, the specific decision to be made provides direction for the evaluation process which consists of the three steps listed. It should be clear that every decision maker evaluates whether formally or informally. That is, anytime s/he must make a decision s/he will form questions and will seek answers to these questions by utilizing some strategy to gather information--even if it is a very loose and informal strategy.

When someone called an "evaluator" is called in to help with this process her/his function is to assist the decision maker(s), the evaluator becomes in essence an extension of the decision makers mental process. The evaluator helps the decision maker(s) by:

.delineating--helping the decision maker determine what questions should be asked and what information is required.

.obtaining--determining what data gathering strategy to use and by using this strategy to collect data. The strat-

egy will sometimes involve the collection of "new" data while at other times it simply requires compilation of data from existing sources.

.providing--by feeding data back to the decision maker(s) and helping the decision maker(s) utilize the data to make the necessary judgements.

The Phi Delta Kappa Committee on Evaluation (Stufflebeam et al., 1971) has summed up the relationship between decision maker and evaluator under this definition as follows:

> Although there is a close relationship between evaluation and decision making, appropriate lines of demarcation must be drawn between the two roles. Under special circumstances, the decision maker might perform many of his own evaluation tasks of delineating, obtaining, and using evaluative information. This is good to the extent that he can meet the criteria for adequate evaluation. However, the evaluator should by no means assume responsibility for roles inherently those of the decision maker. The evaluator should neither make nor implement program decisions. The essence of his role is to provide information to enable the decision maker to perform these tasks (p. 93).

Summary

I believe that educational evaluation as practiced in the majority of instances has failed to provide practitioners with the information they have needed to improve education.

Much of the blame for this failure of evaluation to provide practitioners with necessary information can be laid squarely on the failure of evaluators to actively involve decision makers at every step of the evaluation process. As a result of these failures, practitioners, especially those involved in developing alternative educational programs, have developed a skeptical (maybe even hostile) view of evaluation.

SEA educators believe that newly emerging alternative educational programs can benefit greatly from formative evaluation information.

However, the evaluators providing the services must be credible in the eyes of the practitioners involved if the efforts are to succeed. To overcome past negativism toward evaluation and to develop credibility, evaluators must be aware of and avoid past evaluation shortcomings. The present paper has outlined many of the more prevalent past evaluation shortcomings and has discussed ways in which the SEA evaluation team believes they may reasonably be avoided.

Evaluation should be considered as an integral part of the decision making process. The specific decision to be made gives direction to the evaluation effort and evaluation provides information to assist in making the decision but it does not make the decision. The evaluator helps the decision maker delineate the information to be gathered, obtains the information, and, finally, provides the information to the decision makers.

References

Cronbach, L. J. Course improvement through evaluation. The Record, 1963, 64, 672-683. Reprinted in P. A. Taylor, & D. M. Cowley, Readings in Curriculum Evaluation. Dubuque, Iowa: Wm. C. Brown. Pp. 11-19.

Gephart, W. J. Three definitions of evaluation. SRIS Quarterly, 1973, 6(2), 4-7.

Hickey, M. Evaluating alternative schools. Changing Schools, 1972, Position Paper No. 1.

Lange, R. R., & Speiss, M. An Empirical Study of four providvariables effecting utilization of evaluation. Paper presented at the 1974 AERA Annual Meeting, Chicago, April, 1974.

Rawitsch, D. Study of participation in governance by representative groups in Southeast Alternatives. Unpublished manuscript, June, 1974, available in ERIC.

Rosen, D. New evaluation for new schools. Changing Schools, 1973, (9), 3-15.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. M. Educational evaluation and decision making. Bloomington: Phi Delta Kappa, 1971.